IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, January 2026
Impact Factor: 7.67

Handwritten Text Digitizer

Dr. M. A. Pradhan, Kshirsagar Prathamesh Rahul, Chavan Aditya Kamlesh, Patel Touhid Sabir
Department of Computer Engineering
All India Shri Shivaji Memorial Society College of Engineering, Pune, Maharashtra
mapradhan@aissmscoe.com, prathmeshkshirsagar2003@gmail.com
aditya.chavan1504@gmail.com, touhidpatel06@gmail.com

Abstract: *Handwritten text recognition remains a difficult problem due to variations in writing styles, stroke patterns, noise, and inconsistencies in real-world documents. Conventional Optical Character Recognition (OCR) systems rely on convolutional and recurrent networks, which show limitations when dealing with unconstrained handwriting. This paper presents a handwritten text digitization system based on TrOCR, a fully Transformer-driven OCR architecture using a Vision Transformer encoder and a pretrained text Transformer decoder. The system operates end-to-end without convolutional backbones, recurrent decoding, or external language models. A complete methodology is proposed, including preprocessing, patch embedding, autoregressive decoding, and text generation. Experimental evaluation demonstrates the effectiveness of TrOCR for handwritten digitization with improved accuracy over conventional baselines. The system is suitable for document automation, archival digitization, intelligent extraction, and data processing applications*

Keywords: Handwritten Text Recognition, TrOCR, OCR, Transformer, Vision Transformer, Deep Learning, Digitization

DOI: 10.48175/IJARSCT-30865